Creation of a Corpus for Evidence Based Summarisation
Diego Mollá and María Elena Santiago-Martínez
Department of Computing, Macquarie University

Abstract
Evidence Based Medicine (EBM) recommends physicians to incorporate published evidence when providing care for their patients. Several summarisation systems have been proposed to help the physician. However, there is no corpus available to develop and compare such systems. Our corpus is developed for this purpose.

Source: Journal of Family Practice

Our XML Version

How it Was Obtained

Clinical Inquiry: From the source title.
Answer Parts (“snips”): From the answer box.
Evidence Grades: From the answer box.
Answer Justifications (“longs”): Annotated by three annotators (members of the research team), with some overlap to check consistency.
References: PubMed IDs obtained by crowdsourcing (Amazon Mechanical Turk).

Statistics

- 456 questions.
- 1,396 snips.
- 3,036 longs.
- 2,908 distinct references.

Crowdsourcing with Amazon Mechanical Turk

- Each group of 10 references was assigned to 5 Turkers.
- The following automatic checks were made:
  1. Include known PubMed IDs for double-checking.
  2. Check if the ID exists.
  3. Check the percentage of overlapping text between the retrieved title of the annotated ID and the reference text.
  4. Check agreement with other Turkers.
- A random sample of 100 references revealed no errors.